DOCUMENT RESUME

ED 442 875 TM 031 301

AUTHOR Huynh, Huynh
TITLE On Bayesian Rules for Selecting 3PL Binary Items for Criterion-Referenced Interpretations and Creating Booklets for Bookmark Standard Setting.
PUB DATE 2000-04-00
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Criterion Referenced Tests; *Selection; Standards; Test Construction; Test Interpretation; *Test Items
IDENTIFIERS Rasch Model; *Standard Setting; *Three Parameter Model

ABSTRACT

By noting that a Rasch or two parameter logistic (2PL) item belongs to the exponential family of random variables and that the probability density function (pdf) of the correct response (X=1) and the incorrect response (X=0) are symmetric with respect to the vertical line at the item location, it is shown that the conjugate prior for ability is proportional to [I(theta)]superscript alpha where I(theta) is the item information and alpha is a positive constant. When the above prior is applied to a three parameter logistic (3PL) item, the requirement that item selection rules are bound to the traditional formula for correction for random guessing implies that the constant alpha must be 1. Thus, maximum information selection rules for 3PL items are the only rules that are consistent with a Bayesian analysis based on the family of conjugate priors and with the use of the correction-for-guessing formula. (Contains 16 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

TM

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

H. Huynh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 442 875

On Bayesian Rules for Selecting 3PL Binary Items for Criterion-Referenced Interpretations and Creating Booklets for Bookmark Standard Setting

Huynh Huynh University of South Carolina

Abstract

By noting that a Rasch or 2PL item belongs to the exponential family of random variables and that the probability density function (pdf) of the correct response (X=1) and the incorrect response (X=0) are symmetric with respect to the vertical line at the item location, it is shown that the conjugate prior for ability is proportional to $[I(\theta)]^{\alpha}$, where $I(\theta)$ is the item information and α is a positive constant. When the above prior is applied to a 3PL item, the requirement that item selection rules are bound to the traditional formula for correction for random guessing implies that the constant α *must be 1*. Thus, maximum information (MI) selection rules for 3PL items are the *only rules* that are consistent with a Bayesian analysis based on the family of conjugate priors and with the use of the correction-for-guessing formula.

Notes

A paper presented at the annual meeting of American Educational Research Association, New Orleans, April 2000. *The author's address is College of Education, Wardlaw 138, University of South Carolina, Columbia, SC 29208. Email address: huynh-huynh@sc.edu.*

TM031301

BEST COPY AVAILABLE

Introduction

Criterion-referenced (CR) measures have been used extensively in the United States in the last several years. Such measures, according to Glaser (1963) would provide explicit information as to what an individual can do or cannot do on a continuum of achievement. Across the years, procedures have been developed to provide meaningful interpretations of test scores. For tests used in the National Assessment of Educational Progress (NAEP) (Beaton & Allen, 1992), for example, CR interpretation is referred to as *scale anchoring*. Criterion-referencing is also the basic concept that underlines the *CTB Bookmark standard setting procedure* (Lewis & Mitzel, 1995; Lewis, Mitzel, & Green, 1996) and other procedures of similar nature such as the one used in the Maryland School Performance Assessment Program (MSPAP) (Westat, 1993, 1994).

In general, to describe a point on a NAEP achievement continuum (an *anchor point*) through *scale anchoring*, a set of items is first selected based on some specified statistical criteria. A content expert committee is convened to examine the items and then arrive at a general description of the skills and performances that are expected from examinees at the anchor point. Similarly, the CTB Bookmark standard setting process begins with placing all items on the achievement continuum and creating an ordered test form. Judges are then asked to place a bookmark at a place in the ordered test form that represents their cutoff score for the proficiency level under consideration (such as the basic, proficient, and advanced levels used in many state assessment programs.) Once a cutoff score is finalized, a group of judges are asked to look at the items that surround the cutoff and determine the nature of the skills associated with this level of achievement.

Historically, the statistical criteria for selecting items for scale anchoring rely on the probability associated with the correct response at various anchor points. Consider a binary item and let p^+_{i-1} and p^+_i be the proportion of correct responses at two successive anchor points $\theta_{i-1}$and θ_i. Beaton and Allen (1992) indicated that, in some NAEP situations, an item was selected to describe anchor point θ_i if $p^+_i \geq .80$ and $p^+_{i-1} \leq .50$. In other NAEP cases, such as the 1990 mathematics scale anchoring, an item was selected if $p^+_i \geq .65$, $p^+_{i-1} \leq .50$, and $p^+_i - p^+_{i-1} \geq .30$. In more recent years (Allen, Kline, & Zelanak, 1996, p. 265), NAEP scale anchoring has been based only on the rule that specifies that $p^+_i \geq .65$ for binary items with no guessing

and $p_i^+ \geq .74$ for multiple-choice items with four options. The proportion of .74 can be derived from the proportion of .65 by using the traditional rule regarding correction for guessing. In fact, if .65 represents the proportion of examinees who know the item (and therefore answer it correctly), then the proportion of examinees who do not know the item is .35. For a multiple-choice item with four options, traditional correction for guessing stipulates that one-fourth of the latter examinees ($.35 \div 4 = .09$) would guess the item correctly. So for multiple-choice items with four options, the threshold value of .65 is now raised to $.65 + .09 = .74$.

Earlier work on the CTB Bookmark process relies on the statistical rule $p_i^+ = .50$ for placing an item without guessing on the achievement continuum (Lewis & Mitzel, 1995). The latest version of this standard setting process (Lewis *et al.*, 1996) is based on the statistical rule $p_i^+ = 2/3$ or .67. The formula for correction for random guessing is used to adjust these cutoff probabilities for multiple-choice items. For an item with four alternatives, for example, the cutoff probability of .50 is adjusted to $.50 + (.50 \div 4) = .63$. As for the cutoff probability of 2/3 or .67 for an item without guessing, the adjusted probability is reset at $.67 + (.33 \div 4) = .75$.

It may be noted that the NAEP rules for scale anchoring are based largely on practical experience and feasibility, and not on any theoretical consideration. In the context of standard setting, Huynh (1994) points out to the need to know what a student can be expected to do or to know at a given on the achievement. Using the Bock (1972) partition of the Fisher item information of a Rasch binary item to its correct response, Huynh arrived at the selection criterion $p_i^+ > 2/3$ or .67 for a binary item without guessing. Subsequently, Huynh (1998) extended the work on selection rules to three-parameter logistic (3PL) and polytomous items. In the 1998 paper, Huynh started with a Bayesian framework with a prior that is proportional to the item information. It turns out that this Bayesian approach is equivalent to the use of the Bock (1972) partition of the item information to each of its categories. For a 3PL item, the maximum information (MI) rule was found to be $p_i^+ > (2 + c)/3$ where c is the guessing parameter. Huynh called this the "*principal rule.*"

This paper extends the work by Huynh (1998) on item selection rules for scale anchoring and Bookmark construction of ordered test forms. Its major purpose is to further explore the Bayesian method for the same topic. Attention will be focussed on the family of conjugate priors in the exponential family of probability density functions (pdf). It will be shown in the last part

of the paper that, if the family of conjugate priors for Rasch or 2PL items is also used for 3PL items, then the maximum information (MI) principal rule presented in Huynh (1998) is the *only rule* that is consistent with the use of the traditional formula for correction for random guessing.

General Bayesian Framework

The reader may note that this paper deals only with the *characteristics of a given item* (along with its score categories) in a latent trait setting. To frame the problem within a mathematical statistics context, the item is treated synonymously as a random variable and its probability density function (pdf). A random sample from this random variable is assumed to exist. This existence implies the notion of independent *repeated testings on the same item* or *identical items* and is assumed in conceptualizing and using the Fisher information. In addition, the search for an appropriate prior for the item is equivalent to the process of *placing the item at the ability that is most suitable for the item.*

Given the above general remarks, the Bayesian approach used for binary items in this paper can also be brought into an empirical Bayes context. Given an item (that is defined by its parameters in a latent trait setting), two general questions will be asked.

Question 1: To which population of examinees is the item most suitable?

Question 2: Among examinees of this population, what is the typical ability of those who answer the item correctly?

Within an empirical Bayes context, answering the first question amounts to searching for a prior that is suitable for the item. As for the second question, a typical ability is often found among the class of Bayes ability estimates associated with the correct response.

It may be noted that the answer to Question 1 varies with each item. This is due to the fact that *each item is most suitable only for a certain point on the achievement continuum (latent trait).* (This is typically the point where the item information is maximized.) Therefore, it expected that individual items would be assigned different priors in a Bayesian analysis like the one explored in this paper.

Now let τ_1 be the answer to Question 2 and θ be any point of the achievement continuum. We will use the following definition in Huynh (1998, p. 47, Definition 2).

Definition. An examinee with ability θ is said "to be expected to answer the item correctly" if $\theta \geq \tau_1$.

To select an item for describing the anchor point θ_i, it will be assumed that examinees at this point have the skills and knowledge beyond what is expected of the correct response to the item. The above definition will now lead to an item being selected as an anchoring item if $\theta_i \geq \tau_1$. Equivalently, let p_i^+ and p_1 be the proportion of correct responses at θ_i and τ_1. The item selection rule can be stated as follows.

Rule: Select an item if $p_i^+ \geq p_1$.

A Review of MI Item Selection

A major purpose of this paper is to find an appropriate value for p_1. Consider a binary item that follows a three-parameter logistic (3PL) model with traditional parameters a, b, and c. Let $P(\theta)$ be the probability associated with the correct response $X = 1$ and $Q(\theta) = 1 - P(\theta)$ be the probability associated with the incorrect response $X = 0$. The probability $P(\theta)$ is given by the following formula:

$$P(\theta) = c + (1 - c)\frac{\exp[a(\theta - b)]}{1 + \exp[a(\theta - b)]}. \tag{1}$$

For ease of notations, let $K = a^2(1 - c)^2$, $P = P(\theta)$, and $Q = Q(\theta)$. Then the item information is known to be equal to

$$I(\theta) = K(1 - P)(P - c)^2/P. \tag{2}$$

Note that the constant K does not depend on either θ or P.

Following a suggestion by Bock (1972, Equation 24), Huynh (1998) partitions the total item information $I(\theta)$ to each of the two responses $X = 0$ and $X = 1$ according to the probabilities $P(\theta)$ and $Q(\theta)$. More specifically the information assigned to the correct response $X = 1$ is taken as

$$I_1(\theta) = I(\theta)P(\theta) = K(1 - P)(P - c)^2. \tag{3}$$

This function is maximized at the value $\theta_1 = b + (\log 2)/a$, or equivalently, at the value $P = p_1 = (c + 2)/3$. From the value p_1, Huynh (1998, p. 48) stated the following rule for selecting items for scale anchoring.

Maximum Information Rule: Select an item if $p_i^+ \geq (c + 2)/3$.

Thus a binary Rasch or 2PL item without guessing is selected if $p_i^+ \geq 2/3$ (or .67). As for a multiple-choice 3PL item with four options, the constant c may be taken as 1/4 and hence the rule becomes $p_i^+ \geq .75$.

It is noted in the introduction part of this paper that the proportion $p_1 = (c + 2)/3$ for a 3PL item can be derived from the proportion of $2/3$ for a 2PL item by using the traditional rule regarding correction for random guessing. In fact, if $2/3$ represents the proportion of examinees who know the item (and therefore answer it correctly), then the proportion of examinees who do not know the item is $1/3$. For a 3PL item with k options, the constant c may be taken approximately as $1/k$. Let us assume that all examinees who do not know the item will randomly guess at the item. Under this assumption (of random guessing), the proportion of examinees who guess the item correctly is equal to $1/3k$ or $c/3$. Hence, the threshold of $2/3$ for a 2PL item is now raised to $2/3 + c/3$ or $(2 + c)/3$ for a 3PL item.

Bayesian Framework for Rasch or 2PL Items

Conjugate prior and Bayes Score Locations

Consider a binary 2PL item with parameters a and b (and with $c = 0$). Without loss of generality, we will absorb the constant a into the latent trait θ and set $a = 1$ for the rest of this section. The 2PL item now becomes a Rasch binary item with difficulty parameter b. The random variable X that represents the two responses $x = 0, 1$ follows the probability density function (pdf):

$$f_X(x \mid \theta) = \exp[x(\theta - b)] / \{1 + \exp[(\theta - b)]\} . \tag{4}$$

We will now search for the family of conjugate priors for Θ. To do this, let $\mathbf{X} = (x_1, ..., x_n)$ be a random sample from the random variable X and let its likelihood be written as

$$f_{\mathbf{X}}(x_1, ..., x_n \mid \theta) = \exp[(\theta - b) \sum_{i=1}^{n} x_i] / \{1 + \exp[(\theta - b)]\}^n . \tag{5}$$

In the context of psychometric theory, this random sample may be conceptualized as the responses of a given examinee of ability θ from n independent repeated administration of the (same) item or the independent administration of n identical items. The random sample can also be thought as the responses to the (same) item from a random sample of n examinees with identical ability θ. See Hambleton & Swaminathan (1990; p. 27) for a discussion on these interpretations.

From the derivations presented in Bernardo and Smith (1994, pp. 266-267), it follows that the family of conjugate priors pdf for the Rasch binary item will take the general form

$$f_{\Theta}(\theta) = \exp[\alpha(\theta - b)] / \{1 + \exp[(\theta - b)]\}^{\beta} / K(\alpha, \beta) \tag{6}$$

where α and β are any positive constants with $\alpha < \beta$, and $K(\alpha, \beta)$ is a suitable constant that depends on α and β.

It may also be noted that the response X of a Rasch binary item is also a Bernoulli random variable with success probability of

$$p = \exp[(\theta - b)] / \{1 + \exp[(\theta - b)]\}.$$

Hence the family of conjugate priors takes the form of the beta function in the argument p. This beta pdf is equal to

$$f_P(p) = p^{u-1}(1-p)^{v-1} / B(u, v) \tag{7}$$

where u and v are positive constants and $B(u, v)$ is a suitable positive constant that depends on u and v. By taking note of equation (7) and the following partial derivative

$$\frac{\delta p}{\delta \theta} = \exp[(\theta - b)] / \{1 + \exp[(\theta - b)]\}^2 = p(1-p), \tag{8}$$

it follows that the pdf of Θ is given as

$$f_{\Theta}(\theta) = p^u (1-p)^v / B(u, v) \tag{9}$$

or

$$f_{\Theta}(\theta) = K(u, v) \exp[u(\theta - b)] / \{1 + \exp[(\theta - b)]\}^{(u+v)} \tag{10}$$

where $K(u, v)$ is a suitable constant. By taking $\alpha = u$ and $\beta = u + v$, it may be seen that the pdf of Θ in equation (6) is identical to the pdf of Θ in equation (10).

As in Huynh (1998), we will take the modal Bayes estimate τ_x for Θ to be the Bayes score location of the response x. This score location may be computed by setting the derivative (with respect to θ) of $\log P_x(\theta) f_{\Theta}(\theta)$ to zero. The process yields the score location

$$\tau_x = b + \log[(x + \alpha)/(\beta + 1 - x - \alpha)]. \tag{11}$$

Thus the Bayes location of the correct response $x = 1$ is

$$\tau_1 = b + \log[(\alpha + 1)/(\beta - \alpha)]. \tag{12}$$

Taking into account the identity $\log[\exp(z)] = z$, it may be verified that at this Bayes location, the probability of getting the correct response is

$$p_1 = (\alpha + 1)/(\beta + 1). \tag{13}$$

Symmetric prior and item information function

It may be noted that the probabilities $f_X(X = 0 \mid \theta)$ and $f_X(X = 1 \mid \theta)$ of the incorrect $(x = 0)$ and correct response $(x = 1)$ to the Rasch item are symmetric with respect to the (vertical) line $\theta = b$. Therefore, *a priori*, it may make sense to treat these responses "equally" by requiring that the prior of Θ be symmetric with respect to the line $\theta = b$. This condition is equivalent to the requirement that the Bayes score locations of the incorrect and correct responses are symmetric with respect to the item location b. This symmetry requirement is fulfilled when $\beta = 2\alpha$. The symmetric conjugate prior pdf for Θ now takes the special form

$$f_\Theta(\theta) = \left\{\exp(\theta - b)/[1 + \exp(\theta - b)]^2\right\}^\alpha / K(\alpha, 2\alpha). \tag{14}$$

With the item information being

$$I(\theta) = \exp(\theta - b)/[1 + \exp(\theta - b)]^2, \tag{15}$$

the symmetric conjugate prior for Θ may be written as

$$f_\Theta(\theta) = [I(\theta)]^\alpha / K(\alpha, 2\alpha). \tag{16}$$

General Empirical Bayes Rule

With this symmetric prior, the Bayes location of the correct response is

$$\tau_1 = b + \log[(\alpha + 1)/\alpha)]. \tag{17}$$

Taking into account the identity $\log[\exp(z)] = z$, it may be verified that at this Bayes location, the probability of getting the correct response is

$$p_1 = (\alpha + 1)/(2\alpha + 1). \tag{18}$$

Thus the general empirical Bayes rule can be stated as follows.

General Empirical Bayes Rule: Select an item if $p_i^+ > (\alpha + 1)/(2\alpha + 1)$, where α is a positive constant.

With α being positive, the quantity $(\alpha + 1)/(2\alpha + 1)$ is larger than .50. Thus the general empirical Bayes rule specifies that an item is selected if the probability of answering correctly is larger than .50. The remainder of this section addresses two special cases regarding the parameter α.

When $\alpha = \frac{1}{2}$, this prior becomes a member of the family of noninformative priors that were proposed and studied by Jeffreys (1939, 1948, 1961). Methods for constructing Jeffreys' priors (and other noninformative or reference priors) may be found in Berger and Bernardo (1992), Lehmann (1983, p. 241), and Schervish (1995, pp. 121-123). With Jeffreys' prior, the Bayes score location for the correct response is $\tau_1 = b + \log 3$. At this location, the p_1 probability is 3/4 (or .75).

The other special case is for $\alpha = 1$. This corresponds to the prior that was considered by Huynh (1998) as a way to introduce the Bock partition of item information (Bock, 1972, Equation 24) to each of the two responses $x = 0$ and $x = 1$. The Bayes score location for the correct response is $\tau_1 = b + \log 2$ and the p_1 probability is 2/3 (or .67).

Bayesian Framework for 3PL Items

Consider now a 3PL item with parameters a, b, and c and with item information given as

$$I(\theta) = K(1 - P)(P - c)^2/P. \tag{19}$$

As in the case of Rasch or 2PL items, we will consider the family of priors that are proportional to $[I(\theta)]^\alpha$ where α is a positive constant. The Bayes score location for the correct response is the value of θ at which the function $[I(\theta)]^\alpha P$ is maximized. Since P is an increasing function in θ, maximization may be accomplished by finding the value of P at which the derivative (with respect to P) of $\log\{[I(\theta)]^\alpha P\}$ is equal to zero. Thus P is a solution of the equation

$$\alpha[-\frac{1}{1-P} + \frac{1}{P-c} - \frac{1}{P}] + \frac{1}{P} = 0. \tag{20}$$

Algebraic manipulations yields the following quadratic equation for P.

$$(2\alpha + 1)P^2 - (\alpha + c + 1)P + c(1 - \alpha) = 0. \tag{21}$$

The *function* on the left side of this equation is negative at $P = c$ and positive at $P = 1$. Therefore, the above quadratic equation has only one solution between c and 1. Without undue difficulty, it may also verified that this solution maximizes the function $[I(\theta)]^a P$. Thus the Bayes score location for the correct response is the value θ at which the probability of the correct response is equal to

$$p_1^* = \frac{\alpha + c + 1 + [(\alpha + c + 1)^2 - 4c(2\alpha + 1)(1 - \alpha)]^{1/2}}{2(2\alpha + 1)}. \tag{22}$$

As an illustration, consider a 3PL item with $c = .25$. With Jeffreys' prior ($a = .5$), the quadratic equation becomes

$$2P^2 - 1.75P + .125 = 0. \tag{23}$$

This equation yields the solution $P = p_1^* = 0.80$. (The other solution $P = .08$ is smaller than the value of c and, therefore, is not acceptable.) As noted in the last section, for a Rasch or 2PL item, the threshold probability for item selection is $p_1 = .75$. Thus,under Jeffreys' prior, a Rasch or 2PL item is selected for anchor point θ_i if $p_i^\dagger \geq .75$. As for a 3PL item with four options (and with $c = .25$), the selection rule is $p_i^\dagger \geq .80$. It may be verified that the threshold value for $p_i^\dagger$ (namely .80) for a 3PL item (with guessing) *cannot* be deduced from the value .75 for a 2PL item (without guessing) from the formula for correction for random guessing

The next section will investigate the condition under which the formula for correction for guessing can be used to relate these two threshold values for $p_i^\dagger$.

Conditions Under Which Formula for Correction for Random Guessing Can Be Used

Now for a 2PL item, the p_1 probability was found (Equation 18) in the previous section to be

$$p_1 = (\alpha + 1)/(2\alpha + 1). \tag{24}$$

Under correction for random guessing, the value p_1^* (Equation 22) is related to the value p_1 via the formula

$$p_1^* = p_1 + c(1 - p_1). \tag{25}$$

In other word, we have

$$p_1^* = [\alpha(1 + c) + 1]/(2\alpha + 1). \tag{26}$$

Replacing this value of p_1^* in equations (20) or (21) and after some straightforward algebraic manipulations, the following equation will be obtained.

$$c(2\alpha + 1)^2(\alpha - 1) = 0. \tag{27}$$

This equation is satisfied for all values of c if and only if $\alpha = 1$. This value for α corresponds to the prior considered in Huynh (1998) as a precursor to the analysis of score locations of 3PL items based on the Bock partition of the item information.

Summary

This paper extends the work by Huynh (1994, 1998) on the topic of selecting items for scale anchoring and criterion-referenced interpretation. It focuses on a Bayesian analysis based on the family of conjugate priors for Rasch and 2PL items. More specifically, it is be shown that if this family of conjugate priors is also used for 3PL items, then the maximum information (MI) (principal) rule presented in Huynh (1998) is the *only rule* that is consistent with the use of the traditional formula for correction for random guessing.

References

Allen, N. L., Kline, D. L., & Zelenak, C. A. (1996) (Eds.) *The NAEP 1994 Technical Report.* Washington, DC: U.S. Department of Education.

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics, 17*, 191-204.

Berger, J. O., & Bernardo, J. M. (1992). On the development of reference priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 35-60. Oxford: Oxford University Press.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory.* New York: Wiley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika, 37*, 29-51.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Hambleton, R. K., & Swaminathan, H. (1990). *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Huynh, H. (1994). Some technical aspects of standard setting. In *Proceeding of the Joint Conference on Standard Setting for Large-Scale Assessments* (pp.75-93). Washington, DC: National Assessment Governing Board and National Center for Education Statistics.

Huynh, H. (1998). On score locations for binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics, 23*, 35-56.

Jeffreys, H. (1939, 1948, 1961). *The Theory of Probability.* Oxford : Oxford University Press.

Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: John Wiley.

Lewis, D. M., & Mitzel, H. C. (1995). *An item response theory based standard setting procedure.* In D. R. Green (Chair), Some uses of item response theory in standard setting. Symposium presented at the annual meeting of the California Educational Research Association, Lake Tahoe, NV.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A Bookmark approach.* Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.

Schervish, M. J. (1995). *Theory of Statistics.* New York: Spring-Verlag.

Westat (1993. *Establishing proficiency levels and descriptions for the 1992 Maryland School Performance Assessment Program (MSPAP).* Baltimore: Maryland State Department of Education (Author).

Westat (1994). *Establishing proficiency levels and descriptions for the 1993 Maryland School Performance Assessment Program (MSPAP).* Baltimore: Maryland State Department of Education (Author).